Quantum Coding Theorems for Arbitrary Sources, Channels and Entanglement Resources

Garry Bowen and Nilanjana Datta

***Abstract*— The information spectrum approach gives general formulae for optimal rates of various information theoretic protocols, under minimal assumptions on the nature of the sources, channels and entanglement resources involved. This paper culminates in the derivation of the dense coding capacity for a noiseless quantum channel, assisted by arbitrary shared entanglement, using this approach. We also review the currently known coding theorems, and their converses, for protocols such as data compression for arbitrary quantum sources and transmission of classical information through arbitrary quantum channels. In addition, we derive the optimal rate of data compression for a mixed source.**

***Index Terms*— Quantum information, dense coding capacity, quantum data compression, classical capacity, information spectrum.**

I. Introduction

QUANTUM information theory generalizes the ideas of coding and communication to include the nature of the physical system in which information is encoded. The information spectrum approach of Han & Verdu [2], [3] gives general formulae for many operational schemes in information theory. It replaces the idea of typical events (generally called typical sequences) in information theory, with high probability events. The power of this approach lies in the lack of assumptions about the source, channel and entanglement resource.

The quantum information spectrum was defined in terms of quantum states by Hayashi & Nagaoka [6], initially in the context of hypothesis testing, and was used to determine a general expression for the classical capacity of arbitrary quantum channels. The quantum information spectrum extends the idea of high probability events to high probability subspaces of states in a Hilbert space. In the commutative case, the quantum information spectrum simply reduces to its classical counterpart.

In this paper we present a review of coding theorems for quantum data compression and transmission of classical information through a quantum channel. The rate of compression for a mixed source is explicitly derived. A number of new results are also presented, including the dense coding capacity for a noiseless quantum channel, assisted by arbitrary shared entanglement.

This work was supported by the EPSRC (Research Grant GR/S92816).

G. Bowen is with the Centre for Quantum Computation, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK (e-mail: gab30@damtp.cam.ac.uk).

N. Datta is with the Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge CB3 0WA, UK (e-mail: N.Datta@statslab.cam.ac.uk).

II. Preliminaries

Let $\mathcal{B}(\mathcal{H})$ denote the algebra of linear operators acting on a finite–dimensional Hilbert space $\mathcal{H}$ of dimension d. The von Neumann entropy of a state ρ, i.e. a positive operator of unit trace in $\mathcal{B}(\mathcal{H})$, is defined as $S(\rho) = -\mathrm{Tr}\rho\log\rho$. Throughout this paper, we choose the logarithm to base e. We could equally well choose an arbitrary base for the logarithm. This would simply scale the unit of information.

A quantum channel is given by a completely positive trace–preserving (CPTP) map $\Phi : \mathcal{B}(\mathcal{K}) \to \mathcal{B}(\mathcal{H})$, where $\mathcal{K}$ and $\mathcal{H}$ are the input and output Hilbert spaces of the channel.

A. Spectral Projections

The quantum information spectrum approach requires the extensive use of spectral operators. For a self-adjoint operator A written in its spectral decomposition $A = \sum_i \lambda_i |i\rangle\langle i|$ we define the positive spectral projection on A as

$$\{A \geq 0\} = \sum_{\lambda_i \geq 0} |i\rangle\langle i| \tag{1}$$

the projector onto the eigenspace of positive eigenvalues of A. Corresponding definitions apply for the other spectral projections $\{A < 0\}$, $\{A > 0\}$ and $\{A \leq 0\}$. For two operators A and B, we can then define $\{A \geq B\}$ as $\{A - B \geq 0\}$, and similarly for the other ordering relations.

B. Two Important Lemmas

The following key lemmas are used repeatedly in the paper. For their proofs see [1].

Lemma 1: For self-adjoint operators A, B and any positive operator $0 \leq P \leq I$ the inequality

$$\mathrm{Tr}\big[P(A-B)\big] \leq \mathrm{Tr}\big[\{A \geq B\}(A-B)\big] \tag{2}$$

holds.

Lemma 2: For self-adjoint operators A and B, and any completely positive trace-preserving (CPTP) map $\mathcal{T}$ the inequality

$$\mathrm{Tr}\big[\{\mathcal{T}(A) \geq \mathcal{T}(B)\}\mathcal{T}(A-B)\big] \leq \mathrm{Tr}\big[\{A \geq B\}(A-B)\big] \tag{3}$$

holds.

We also make use of the following proposition

Proposition 1: Given a state ρ_n and a self-adjoint operator ω_n, we have

$$\mathrm{Tr}\big[\{\rho_n \geq e^{n\gamma}\omega_n\}\omega_n\big] \leq e^{-n\gamma}. \tag{4}$$

for any real γ.

Proof: We have

$$\mathrm{Tr}\big[\{\rho_n \geq e^{n\gamma}\omega_n\}(\rho_n - e^{n\gamma}\omega_n\big] \geq 0 \tag{5}$$

and hence, by rearranging terms

$$\mathrm{Tr}\big[\{\rho_n \geq e^{n\gamma}\omega_n\}\omega\big] \leq e^{-n\gamma}\mathrm{Tr}\big[\{\rho_n \geq e^{n\gamma}\omega_n\}\rho_n\big] \leq e^{-n\gamma}. \tag{6}$$

where $\mathrm{Tr}\big[\{\rho_n \geq e^{n\gamma}\omega_n\}\rho_n\big] \leq 1$. ∎

C. Quantum Spectral Information Rates

As a generalization of the relative entropy, the spectral divergence allows information theory to include arbitrary sources and channels.

Definition 1: Given the difference operator $\Pi_n(\gamma) = \rho_n - e^{n\gamma}\omega_n$, the quantum spectral sup-(inf-)divergence rates are defined as

$$\overline{D}(\rho\|\omega) = \inf\left\{\gamma : \limsup_{n\to\infty} \mathrm{Tr}\big[\{\Pi_n(\gamma) \geq 0\}\Pi_n(\gamma)\big] = 0\right\} \tag{7}$$

$$\underline{D}(\rho\|\omega) = \sup\left\{\gamma : \liminf_{n\to\infty} \mathrm{Tr}\big[\{\Pi_n(\gamma) \geq 0\}\Pi_n(\gamma)\big] = 1\right\} \tag{8}$$

respectively.

The spectral entropies, conditional spectral entropies, and spectral mutual information rates may all be expressed as a divegrence rate with appropriate substitutions for the sequence of operators $\omega = \{\omega_n\}_{n=1}^{\infty}$. These are

$$\overline{S}(\rho) = -\underline{D}(\rho|I) \tag{9}$$
$$\underline{S}(\rho) = -\overline{D}(\rho|I) \tag{10}$$

and for sequences of bipartite state $\rho^{AB} = \{\rho_n^{AB}\}_{n=1}^{\infty}$,

$$\overline{S}(A|B) = -\underline{D}(\rho^{AB}|I^A \otimes \rho^B) \tag{11}$$
$$\underline{S}(A|B) = -\overline{D}(\rho^{AB}|I^A \otimes \rho^B) \tag{12}$$
$$\overline{S}(A:B) = \overline{D}(\rho^{AB}|\rho^A \otimes \rho^B) \tag{13}$$
$$\underline{S}(A:B) = \underline{D}(\rho^{AB}|\rho^A \otimes \rho^B), \tag{14}$$

giving all the spectral sup(inf)-information rates. Various properties and relationships of these quantities are explored in [1].

III. Data Compression for Arbitrary Quantum Sources

A general quantum source consists of a sequence of density $\rho = \{\rho_n\}_{n=1}^{\infty}$ acting on a corresponding sequence of Hilbert spaces $\mathcal{H} = \{\mathcal{H}_n\}_{n=1}^{\infty}$.

A compression scheme for such a source, ρ, consists of two families of quantum operations $\mathcal{C}_n$ and $\mathcal{D}_n$. Here $\mathcal{C}_n$ denotes the compression operation which takes states in the original Hilbert space $\mathcal{H}_n$ to states in a Hilbert space $\widetilde{\mathcal{H}_n}$ such that dim $\widetilde{\mathcal{H}_n} \leq$ dim $\mathcal{H}_n$. Hence, $\widetilde{\mathcal{H}_n}$ can be regarded as the compressed Hilbert space. The corresponding decompression operation, $\mathcal{D}_n$, takes states in $\widetilde{\mathcal{H}_n}$ to states in the original Hilbert space $\mathcal{H}_n$.

The compression scheme given by the family of combined compression decompression maps $\mathcal{D}_n \circ \mathcal{C}_n$ is said to be *reliable* if the entanglement fidelity $F(\rho_n, \mathcal{D}_n \circ \mathcal{C}_n)$ tends to 1 as $n \to \infty$. Let P_n denote the orthogonal projection onto $\widetilde{\mathcal{H}_n}$. The *rate* of the compression scheme is determined by

$$R = \limsup_{n\to\infty} \frac{1}{n}\log M_n, \tag{15}$$

where $M_n := \mathrm{Tr}P_n = \dim\widetilde{\mathcal{H}_n}$.

The objective is thus to obtain the optimal rate of reliable compression for a given source ρ. Defining the optimal rate $\mathcal{R}$ as the infimum of all reliable rates, leads to the following theorem.

Theorem 1: The quantum spectral sup-entropy rate is optimal. Hence,

$$\mathcal{R} = \overline{S}(\rho) \tag{16}$$

for a given source ρ. Equivalently, (i) if $R > \overline{S}(\rho)$ then there exists a reliable compression scheme of rate R, and (ii) there can be no reliable compression scheme of rate R for $R < \overline{S}(\rho)$.

Proof: [Proof of (i) :] Suppose $R > \overline{S}(\rho)$. Consider the compression operation, $\mathcal{C}_n$, defined by its action on any state $\sigma_n \in \mathcal{B}(\mathcal{H}_n)$ as follows:

$$\mathcal{C}_n(\sigma_n) := P_n\sigma_n P_n + \sum_k A_k\sigma_n A_k^{\dagger}, \tag{17}$$

where (a) P_n, the compression projection, i.e. the orthogonal projection onto the compressed Hilbert space $\widetilde{\mathcal{H}_n}$, is given by

$$P_n := \{\rho_n \geq e^{-n\gamma}I_n\}, \tag{18}$$

and (b) $A_k := |\chi_0\rangle\langle k|$, with $|\chi_0\rangle$ being a fixed pure state in $\widetilde{\mathcal{H}_n}$ and $\{|k\rangle\}$ being an orthonormal basis for the orthocomplement of $\widetilde{\mathcal{H}_n}$. Equivalently,

$$\mathcal{C}_n(\sigma_n) := P_n\sigma_n P_n + \mathrm{Tr}\big((I_n - P_n)\sigma_n\big)|\chi_0\rangle\langle\chi_0|. \tag{19}$$

The corresponding decoding operation $\mathcal{D}_n$ is defined to be the identity on $\widetilde{\mathcal{H}_n}$.

If $\{C_n^j\}$ and $\{D_n^k\}$ denote finite sets of Kraus operators of the quantum operations $\mathcal{C}_n$ and $\mathcal{D}_n$ respectively, then

$$F_n := F(\rho_n, \mathcal{D}_n \circ \mathcal{C}_n) = \sum_{jk} |\mathrm{Tr}\big(D_n^k C_n^j \rho_n)|^2. \tag{20}$$

and hence the entanglement fidelity is given by

$$\begin{aligned} F(\rho_n, \mathcal{D}_n \circ \mathcal{C}_n) &= |\mathrm{Tr}(P_n\rho_n)|^2 + \sum_k |\mathrm{Tr}(A_k\rho_n)|^2 \\ &\geq |\mathrm{Tr}(P_n\rho_n)|^2 \\ &\geq |\mathrm{Tr}\big[P_n(\rho_n - e^{-n\gamma}I_n)\big]|^2 \\ &= |\mathrm{Tr}\big[\{\rho_n \geq e^{-n\gamma}I_n\}(\rho_n - e^{-n\gamma}I_n)\big]|^2 \\ &= |\mathrm{Tr}\big[\{\Pi_n(\gamma) \geq 0\}\Pi_n(\gamma)\big]|^2, \end{aligned} \tag{21}$$

where $\Pi_n(\gamma) = \rho_n - e^{n\gamma}I_n$. From the definitions in (9) and (8) it follows that the RHS of (21) tends to 1 as $n \to \infty$, for any $\gamma > \overline{S}(\rho)$.

Utilizing Proposition 1, the dimension of the compression projections P_n is bounded for each n by

$$\mathrm{Tr}P_n = \mathrm{Tr}\big[\{\rho_n \geq e^{-n\gamma}I_n\}\big] \leq e^{n\gamma} = e^{n(\overline{S}(\rho)+\delta)} \tag{22}$$

for $\delta > 0$. Since this is true for all $\delta > 0$ we have $\mathcal{R} \leq \overline{S}(\rho)$.

[Proof of (ii) (Weak Converse):] Suppose $R < \overline{S}(\rho)$. Without loss of generality, assume that $\mathcal{C}_n$ maps states in $\mathcal{H}_n$ to states in an M_n-dimensional Hilbert space $\widetilde{\mathcal{H}_n}$, with $M_n = \lfloor e^{nR} \rfloor$. Hence, if P_n is the orthogonal projection onto $\widetilde{\mathcal{H}_n}$ then $\mathrm{Tr}[P_n] = M_n \leq e^{nR}$.

Let $\{C_n^j\}$ and $\{D_n^k\}$ denote finite sets of Kraus operators for the quantum operations $\mathcal{C}_n$ and $\mathcal{D}_n$ respectively. Obviously, $P_n C_n^j = C_n^j$. Further, let Q_n^k be the orthogonal projection onto the subspace to which $\widetilde{\mathcal{H}_n}$ is mapped to by D_n^k. Then $D_n^k C_n^j = D_n^k P_n C_n^j = Q_n^k D_n^k P_n C_n^j = Q_n^k D_n^k C_n^j$. Moreover, $\mathrm{Tr}[Q_n^k] \leq \mathrm{Tr}[P_n]$ since $\mathcal{D}_n$ is a CPTP map.

The entanglement fidelity can be expressed as

$$\begin{aligned} F_n &= \sum_{jk} |\mathrm{Tr}\big(D_n^k C_n^j \rho_n\big)|^2 \\ &= \sum_{jk} |\mathrm{Tr}\big(Q_n^k D_n^k C_n^j \rho_n\big)|^2 \\ &= \sum_{jk} |\mathrm{Tr}\big[(D_n^k C_n^j \sqrt{\rho_n})(\sqrt{\rho_n} Q_n^k)\big]|^2 \\ &\leq \sum_{jk} \mathrm{Tr}\big[Q_n^k \rho_n Q_n^k\big] \cdot \mathrm{Tr}\big[D_n^k C_n^j \rho_n C_n^{j\dagger} D_n^{k\dagger}\big] \qquad (23) \\ &\leq \mathrm{Tr}\big[P_n \rho_n\big] \qquad (24) \\ &\leq \mathrm{Tr}\big[\{\rho_n \geq e^{-n\gamma} I_n\}(\rho_n - e^{-n\gamma} I_n)\big] \\ &\quad + e^{-n\gamma} \mathrm{Tr} P_n \qquad (25) \end{aligned}$$

To arrive at (23), we have made use of the Cauchy Schwarz inequality for the Hilbert-Schmidt inner product, specifically $|\mathrm{Tr}(A^\dagger B)|^2 \leq \mathrm{Tr}(A^\dagger A) \cdot \mathrm{Tr}(B^\dagger B)$. The inequality in (24) uses the inequality $\mathrm{Tr} Q_n^k \leq \mathrm{Tr} P_n$, and the fact that $\mathcal{C}_n$ and $\mathcal{D}_n$ are trace preserving maps. The final inequality in (25) follows from Lemma 1.

Using the fact that $\mathrm{Tr} P_n \leq e^{nR}$, we have

$$F_n \leq \mathrm{Tr}\big[\{\rho_n \geq e^{-n\gamma} I_n\}(\rho_n - e^{-n\gamma} I_n)\big] + e^{-n(\gamma - R)}. \qquad (26)$$

Choosing a number γ and $\delta > 0$ such that $R = \gamma + \delta < \overline{S}(\rho)$, the second term on RHS of (26) tends to zero as $n \to \infty$. However, since $\gamma < \overline{S}(\rho)$ the first term on RHS of (26) does not converge to 1 as $n \to \infty$. Hence, the fidelity does not converge to 1 in the limit as $n \to \infty$ and the compression scheme is not reliable. ∎

The proof of the weak converse above shows that for $R < \overline{S}(\rho)$ the entanglement fidelity cannot approach unity, and hence any compression scheme will give an error with non-zero probability. To determine the rate at which the probability of error converges to 1 for any compression protocol we can equivalently determine the supremum of the rates for which the asymptotic limit of the entanglement fidelity goes to zero. Here we prove the result for the *strong converse* rate denoted by $\mathcal{R}^*$.

Theorem 2: Coding a source ρ at a rate less than the quantum inf-spectral entropy rate gives an error with probability equal to one. That is

$$R < \underline{S}(\rho) \implies \lim_{n\to\infty} F_n = 0. \qquad (27)$$

or, equivalently $\mathcal{R}^* = \underline{S}(\rho)$.

Proof: From (25) we can immediately see that for rates $R < \underline{S}(\rho)$ choosing a $\gamma = R + \delta < \underline{S}(\rho)$ we obtain

$$\lim_{n\to\infty} F_n = 0 \qquad (28)$$

and the compression scheme fails with probability approaching 1 as $n \to \infty$. ∎

A. Relationship to the von Neumann Entropy

For any quantum information source ρ, the quantum spectral sup- and inf- information rates are related to the von Neumann entropy in the following manner.

Lemma 3: The sup-information and inf-information rates are related to the von Neumann entropy by

$$\underline{S}(\rho) \leq \liminf_{n\to\infty} \frac{1}{n} S(\rho_n) \leq \limsup_{n\to\infty} \frac{1}{n} S(\rho_n) \leq \overline{S}(\rho) \qquad (29)$$

for any source ρ.

Proof: Let $\{\lambda_n^i\}$ denote the set of eigenvalues of state ρ_n. For the first inequality we have

$$\begin{aligned} \frac{1}{n} S(\rho_n) &= -\frac{1}{n} \mathrm{Tr}\big[\rho_n \log \rho_n\big] \\ &= -\frac{1}{n} \sum_i \lambda_n^i \log \lambda_n^i \\ &\geq -\frac{1}{n} \sum_{\lambda_n^i < e^{-n(\underline{S}(\rho)-\delta)}} \lambda_n^i \log \lambda_n^i \\ &\geq -\frac{1}{n} \mathrm{Tr}\big[\{\rho_n < e^{-n(\underline{S}(\rho)-\delta)}\}\rho_n\big] \log e^{-n(\underline{S}(\rho)-\delta)} \\ &= (\underline{S}(\rho) - \delta) \mathrm{Tr}\big[\{\rho_n < e^{-n(\underline{S}(\rho)-\delta)}\}\rho_n\big] \qquad (30) \end{aligned}$$

and from the definition of $\underline{S}(\rho)$ we have $\lim_{n\to\infty} \mathrm{Tr}\big[\{\rho_n \leq e^{-n(\underline{S}(\rho)-\delta)}\}\rho_n\big] = 1$, and this is true for all $\delta > 0$, implying

$$\underline{S}(\rho) \leq \liminf_{n\to\infty} \frac{1}{n} S(\rho_n) \qquad (31)$$

Similarly, we have

$$\begin{aligned} \frac{1}{n} S(\rho_n) &= -\frac{1}{n} \sum_{\lambda_n^i \geq e^{-n(\overline{S}(\rho)+\delta)}} \lambda_n^i \log \lambda_n^i \\ &\qquad - \frac{1}{n} \sum_{\lambda_n^i < e^{-n(\overline{S}(\rho)+\delta)}} \lambda_n^i \log \lambda_n^i \\ &\leq (\overline{S}(\rho) + \delta) \mathrm{Tr}\big[\{\rho_n \geq e^{-n(\overline{S}(\rho)+\delta)}\}\rho_n\big] \\ &\quad - \frac{1}{n} \mathrm{Tr}\big[Q_n \rho_n \log\, \rho_n\big], \qquad (32) \end{aligned}$$

where $Q_n := \{\rho_n < e^{-n(\overline{S}(\rho)+\delta)}\}$. Let $W_n := Q_n \rho_n Q_n$ and define the normalized state $\widehat{W}_n := W_n / (\mathrm{Tr} W_n)$. Hence,

$$\begin{aligned} \frac{1}{n} S(\rho_n) &\leq (\overline{S}(\rho) + \delta) \mathrm{Tr}\big[\{\rho_n \geq e^{-n(\overline{S}(\rho)+\delta)}\}\rho_n\big] \\ &\quad - \frac{1}{n} \mathrm{Tr} W_n \big(\log\, W_n + \log \mathrm{Tr} W_n - \log \mathrm{Tr} W_n\big) \\ &= (\overline{S}(\rho) + \delta) \mathrm{Tr}\big[\{\rho_n \geq e^{-n(\overline{S}(\rho)+\delta)}\}\rho_n\big] \\ &\quad - \frac{1}{n} \mathrm{Tr} W_n S(\widehat{W}_n) - \frac{1}{n} H(\mathrm{Tr} W_n), \\ &\leq (\overline{S}(\rho) + \delta) \mathrm{Tr}\big[\{\rho_n \geq e^{-n(\overline{S}(\rho)+\delta)}\}\rho_n\big] \\ &\quad + \frac{1}{n} \log d_n\, \mathrm{Tr} W_n - \frac{1}{n} H(\mathrm{Tr} W_n) \qquad (33) \end{aligned}$$

In the above, $H(\cdot)$ denotes the Shannon entropy. and $d_n = \dim \mathcal{H}_n$. Since $\lim_{n\to\infty} \mathrm{Tr} W_n = \lim_{n\to\infty} \mathrm{Tr}\big[\{\rho_n < e^{-n(\overline{S}(\rho)+\delta)}\}\rho_n\big] = 0$, the last term vanishes in this limit. The second term also vanishes under the assumption that for all n

$$\frac{1}{n} \log d_n < \beta \tag{34}$$

for some $\beta < +\infty$. Moreover, since $\lim_{n\to\infty} \mathrm{Tr}\big[\{\rho_n \geq e^{-n(\overline{S}(\rho)+\delta)}\}\rho_n\big] = 1$, we have

$$\limsup_{n\to\infty} \frac{1}{n} S(\rho_n) \leq \overline{S}(\rho). \tag{35}$$

The remaining inequality follows from the definition of $\liminf$ and $\limsup$. ■

B. Mixed Sources

Given two sources $\sigma = \{\sigma_n\}_{n=1}^{\infty}$ and $\omega = \{\omega_n\}_{n=1}^{\infty}$, we define the mixed source $\rho = \{\rho_n\}_{n=1}^{\infty}$ to be the source for which

$$\rho_n = t\sigma_n + (1-t)\omega_n \tag{36}$$

for $t \in (0,1)$.

Theorem 3: For the mixed source ρ the optimal rate $\mathcal{R}$ is given by

$$\mathcal{R} = \max\left[\overline{S}(\sigma), \overline{S}(\omega)\right], \tag{37}$$

the maximum of the rates for either source σ or ω.

Proof: Let $\mathrm{Tr}\big[\Pi_n(\gamma)\big] = \mathrm{Tr}\big[\{\rho_n \geq e^{-n\gamma} I_n\}(\rho_n - e^{-n\gamma} I_n)\big]$, then from the linearity of the trace operation, we have

$$\begin{aligned}
\mathrm{Tr}\big[\Pi_n(\gamma)\big] &= t\,\mathrm{Tr}\big[\{\rho_n \geq e^{-n\gamma} I_n\}(\omega_n - e^{-n\gamma} I_n)\big] \\
&\quad + (1-t)\,\mathrm{Tr}\big[\{\rho_n \geq e^{-n\gamma} I_n\}(\sigma_n - e^{-n\gamma} I_n)\big] \\
&\leq t\,\mathrm{Tr}\big[\{\omega_n \geq e^{-n\gamma} I_n\}(\omega_n - e^{-n\gamma} I_n)\big] \\
&\quad + (1-t)\,\mathrm{Tr}\big[\{\sigma_n \geq e^{-n\gamma} I_n\}(\sigma_n - e^{-n\gamma} I_n)\big]
\end{aligned} \tag{38}$$

where the inequality follows from Lemma 1. Hence for any $\gamma = \overline{S}(\rho) + \delta$, the limit of the LHS goes to one, and hence both of the traces on the RHS must also approach one in the limit. This then implies that

$$\overline{S}(\rho) \geq \max\left[\overline{S}(\sigma), \overline{S}(\omega)\right] \tag{39}$$

as δ is arbitrary.

To prove the reverse inequality we explicitly construct a sequence of projection operators. For each $\alpha > 0$ we utilize the projections $P_n^0 := \{\sigma_n \geq e^{-n\alpha} I_n\}$ and $Q_n := \{\omega_n \geq e^{-n\alpha} I_n\}$. Let Q_n have the spectral projection $Q_n = \sum_{i=1}^{K} |i\rangle\langle i|$, with $K = \mathrm{Tr} Q_n$. Starting with P_n^0, we define a sequence of projection operators P_n^i, $i = 1, \ldots, K$, iteratively, as follows. For each i, if $|i\rangle$ lies in the subspace onto which P_n^{i-1} projects, then we set $P_n^i = P_n^{i-1}$. Otherwise, we take the component of $|i\rangle$ orthogonal to this subspace, say $|i^{\perp}\rangle$, and let $P_n^i = P_n^{i-1} \oplus |i^{\perp}\rangle\langle i^{\perp}|$.

From Lemma 1 it then follows that

$$\begin{aligned}
\mathrm{Tr}\big[\Pi_n(\gamma)\big] &\geq \mathrm{Tr}\big[P_n^K(\rho_n - e^{-n\gamma} I_n)\big] \\
&= t\,\mathrm{Tr}\big[P_n^K(\omega_n - e^{-n\gamma} I_n)\big] \\
&\quad + (1-t)\,\mathrm{Tr}\big[P_n^K(\sigma_n - e^{-n\gamma} I_n)\big] \\
&\geq t\,\mathrm{Tr}\big[\{\omega_n \geq e^{-n\alpha} I_n\}\omega_n\big] \\
&\quad + (1-t)\,\mathrm{Tr}\big[\{\sigma_n \geq e^{-n\alpha} I_n\}\sigma_n\big] \\
&\quad - e^{-n\gamma}\mathrm{Tr}\big[P_n^K\big] \\
&\geq t\,\mathrm{Tr}\big[\{\omega_n \geq e^{-n\alpha} I_n\}\omega_n\big] \\
&\quad + (1-t)\,\mathrm{Tr}\big[\{\sigma_n \geq e^{-n\alpha} I_n\}\sigma_n\big] \\
&\quad - 2e^{-n(\gamma-\alpha)}
\end{aligned} \tag{40}$$

where $\mathrm{Tr}\big[P_n^K\big] \leq 2e^{n\alpha}$, as the rank of the projector cannot be greater than the sum of the ranks of the projectors P_n^0 and Q_n. For every $\delta > 0$ and $\alpha = \max\left[\overline{S}(\sigma), \overline{S}(\omega)\right] + \delta$, the limit of the sum of first two terms on the RHS goes to 1. By choosing $\gamma = \alpha + \delta$ this implies both the RHS and LHS converge to 1 and hence that

$$\overline{S}(\rho) \leq \max\left[\overline{S}(\sigma), \overline{S}(\omega)\right].$$

as δ is arbitrary. ■

Corollary 1: The strong converse is given by

$$\mathcal{R}^* = \min\left[\underline{S}(\sigma), \underline{S}(\omega)\right] \tag{41}$$

for any mixed source $\rho_n = t\sigma_n + (1-t)\omega_n$, for $t \in (0,1)$.

Proof: Choosing γ and α such that the RHS and LHS of (38) and (40) go to zero, respectively, gives the required inequalities. ■

A source obeys the strong converse property only if

$$\underline{S}(\rho) = \lim_{n\to\infty} \frac{1}{n} S(\rho_n) = \overline{S}(\rho) \tag{42}$$

Note that mixed sources do not obey the strong converse property if $\max\left[\overline{S}(\sigma), \overline{S}(\omega)\right] > \min\left[\underline{S}(\sigma), \underline{S}(\omega)\right]$. This can easily be shown to hold for mixtures of stationary memoryless sources with different entropies $S(\sigma) > S(\omega)$.

IV. Classical Capacity of an Arbitrary Quantum Channel

In this section we obtain the classical capacity of a sequence of arbitrary quantum channels in terms of the inf-spectral mutual information rate of bipartite separable states shared through the channel.

Let $\{\mathcal{K}_Q^{(n)}\}_{n=1}^{\infty}$ and $\{\mathcal{H}_Q^{(n)}\}_{n=1}^{\infty}$ be two sequences of Hilbert spaces, and let $\Lambda = \{\Lambda_n^Q\}_{n=1}^{\infty}$ be a sequence of quantum channels such that, for each n,

$$\Lambda_n^Q : \mathcal{B}(\mathcal{K}_Q^{(n)}) \mapsto \mathcal{B}(\mathcal{H}_Q^{(n)}).$$

Here $\mathcal{K}_Q^{(n)}$ denotes the Hilbert space at the input of the channel Λ_n^Q, whereas $\mathcal{H}_Q^{(n)}$ denotes the Hilbert space at its output.

Consider the following scenario. Suppose Alice has a set of messages, labelled by the elements of the set $\mathcal{M} = \{1, 2, \ldots, M_n\}$, which she would like to communicate to Bob, using the quantum channel Λ_n^Q. To do this, she encodes each message into a quantum state of a physical system with Hilbert space $\mathcal{K}_Q^{(n)}$ and sends this state to Bob through the quantum channel. In order to infer the message that Alice communicated to him, Bob makes a measurement (described by POVM elements) on the state that he receives. The encoding and decoding operations together define a quantum error correcting code (QECC). More precisely, a code $\mathcal{C}^{(n)}$ of size M_n is given by a sequence $\{\rho_n^i, E_n^i\}_{n=1}^{M_n}$ where each ρ_n^i is a state in $\mathcal{B}(\mathcal{K}_Q^{(n)})$ and each E_n^i is a positive operator acting in $\mathcal{H}_Q^{(n)}$, such that $\sum_{i=1}^{M_n} E_n^i \leq I_n$. Defining $E_n^0 = I_n - \sum_{i=1}^{M_n} E_n^i$, yields a resolution of identity in $\mathcal{H}_Q^{(n)}$. Hence, $\{E_n^i\}_{i=0}^{M_n}$ defines a POVM. An output $i \geq 1$ would lead to the inference that the state ρ_n^i was transmitted through the channel Λ_n^Q, whereas the output 0 is interpreted as a failure of any inference. In other words, a code $\mathcal{C}^{(n)}$ is given by a triple (M_n, ϕ_n, E_n), where ϕ_n is the encoder, i.e., $\phi_n(i) = \rho_n^i$ for $i \in \{1, 2, \ldots, 2^{nR}\}$, and $E_n = \{E_n^i\}_{i=1}^{M_n}$ is the decoder. The rate of the code is given by $(1/n)\log M_n$. The average probability of error for such a code $\mathcal{C}^{(n)}$ is given by

$$P_e(\mathcal{C}^{(n)}) := \frac{1}{M_n} \sum_{i=1}^{M_n} \left(1 - \mathrm{Tr}(\sigma_n^i E_n^i)\right), \tag{43}$$

σ_n^i being the output of the channel when the input is the i^{th} codeword ρ_n^i. A quantity $R \in \mathbf{R}$ is said to be an *achievable rate* if there exists an $N \in \mathbf{N}$ such that for all $n \geq N$, there exits a sequence of codes $\{\mathcal{C}^{(n)}\}_{n=1}^{\infty}$ with $M_n \geq e^{nR}$, and $P_e(\mathcal{C}^{(n)}) \to 0$ as $n \to \infty$.

The capacity of Λ is defined as

$$C(\Lambda) := \sup R, \tag{44}$$

where R is an achievable rate.

Theorem 4: The classical capacity of a sequence of channels $\Lambda = \{\Lambda_n^Q\}_{n=1}^{\infty}$ is given by

$$C(\Lambda) = \max_{\rho^{AQ} \in \mathcal{S}} \underline{S}(A : \Lambda Q) \tag{45}$$

where (i) $\mathcal{S}$ is the set of sequences of separable states in $\mathcal{B}(\mathcal{H}_{AQ})$, with $\mathcal{H}_{AQ}$ being a sequence of Hilbert spaces $\mathcal{H}_{AQ} := \{\mathcal{H}_A^{(n)} \otimes \mathcal{K}_Q^{(n)}\}_{n=1}^{\infty}$, and (ii) $\underline{S}(A : \Lambda Q)$ is the *inf-spectral mutual information rate* of a sequence of separable density matrices $\{\rho_n^{A\Lambda Q}\}_{n=1}^{\infty}$.

Consider an arbitrary set $\mathcal{X}^{(n)}$ of indices and define a separable state

$$\rho_n^{AQ} := \sum_{x \in \mathcal{X}^{(n)}} p_n^x \rho_{n,x}^A \otimes \rho_{n,x}^Q,$$

acting in a Hilbert space $\mathcal{H}_A^{(n)} \otimes \mathcal{K}_Q^{(n)}$. The set of codewords that Alice uses, to transmit her messages to Bob, is a finite subset of the set

$$\{\rho_{n,x}^Q \,:\, x \in \mathcal{X}^{(n)}\}.$$

The state ρ_n^{AQ} can be purified to the state

$$\rho_n^{AA'Q} := \sum_{x \in \mathcal{X}^{(n)}} p_n^x |x\rangle\langle x|^{AA'} \otimes \rho_{n,x}^Q$$

in $\mathcal{B}(\mathcal{H}_A^{(n)} \otimes \mathcal{H}_{A'}^{(n)} \otimes \mathcal{K}_Q^{(n)})$. Let B denote the bipartite system with Hilbert space $\mathcal{H}_A^{(n)} \otimes \mathcal{H}_{A'}^{(n)}$ (and thus replace the superscript AA' by B). Let Q denote system with Hilbert space $\mathcal{K}_Q^{(n)}$. A state ρ_n^{BQ} of the form given by (46) is referred to as a *classical–quantum state*[1]. If X is a random variable with probability mass function $\{p_n^x : x \in \mathcal{X}^{(n)}\}$, then the state of the quantum system Q is correlated with the values taken by the classical index X. The state ρ_n^{BQ} therefore represents the preparation of quantum states $\rho_{n,x}^Q$ corresponding to classical indices $x \in \mathcal{X}^{(n)}$, according to the apriori distribution $\{p_n^x\}$.

The action of the channel Λ_n^Q on the system Q yields the state

$$\begin{aligned} \rho_n^{B\Lambda Q} &:= \left(\mathrm{id}_B \otimes \Lambda_n^Q\right)\left(\rho_n^{BQ}\right) \\ &= \sum_{x \in \mathcal{X}^{(n)}} p_n^x |x\rangle\langle x|^B \otimes \Lambda_n^Q\left(\rho_{n,x}^Q\right) \\ &:= \sum_{x \in \mathcal{X}^{(n)}} p_n^x |x\rangle\langle x|^B \otimes \rho_{n,x}^{\Lambda Q}. \end{aligned} \tag{46}$$

Here the superscript ΛQ is used to denote the system Q after the action of the channel on it.

For the sequence of classical-quantum states $\{\rho_n^{B\Lambda Q}\}$ the *inf-spectral mutual information rate* is given by

$$\underline{S}(B : \Lambda Q) = \sup\left\{\gamma : \lim_{n\to\infty} \mathrm{Tr}\left[\{\Pi_n(\gamma) \geq 0\}\Pi_n(\gamma)\right] = 1\right\} \tag{47}$$

where $\Pi_n(\gamma) := \rho_n^{B\Lambda Q} - \rho_n^B \otimes \rho_n^{\Lambda Q}$, and ρ_n^B, $\rho_n^{\Lambda Q}$ are the reduced density matrices of the systems B and ΛQ respectively.

The proof of the Theorem 4 relies on the following lemma proved in [6].

Lemma 4: For any $n \in \mathbf{N}$, $M \in \mathbf{N}$, and $\gamma \in \mathbf{R}$, given a probability distribution $\{p_n^x\}$ on $\mathcal{X}^{(n)}$, there exists a code $\mathcal{C}^{(n)}$ of size $|\mathcal{C}^{(n)}| = M$, whose average probability of error satisfies the following bound:

$$\begin{aligned} P_e(\mathcal{C}^{(n)}) \leq\ & 2 \sum_{x \in \mathcal{X}^{(n)}} p_n^x \mathrm{Tr}\left[\{\rho_{n,x}^{\Lambda Q} - e^{n\gamma}\rho_n^{\Lambda Q} \leq 0\}\rho_{n,x}^{\Lambda Q}\right] \\ & + 4e^{-n\gamma}M, \end{aligned} \tag{48}$$

where

$$\rho_n^{\Lambda Q} := \sum_{x \in \mathcal{X}^{(n)}} p_n^x \rho_{n,x}^{\Lambda Q}.$$

Proof of Theorem 4 We shall first prove that for any rate $0 < R < \underline{S}(B : \Lambda Q)$, the average probability of error $P_e(\mathcal{C}^{(n)}$ vanishes asymptotically. Here $\underline{S}(B : \Lambda Q)$ denotes the inf-spectral mutual information rate for a sequence of *classical-quantum states* $\{\rho_n^{B\Lambda Q}\}_{n=1}^{\infty}$ and is given by (47).

Computing the reduced density matrices of the bipartite state $\rho_n^{B\Lambda Q}$ (defined by (46)) yields

$$\rho_n^B \otimes \rho_n^{\Lambda Q} = \left(\sum_x p_n^x |x\rangle\langle x|^B\right) \otimes \rho_n^{\Lambda Q}, \tag{49}$$

where $\rho_n^{\Lambda Q} := \sum_x p_n^x \rho_{n,x}^{\Lambda Q}$. The difference operator $\Pi_n(\gamma)$ appearing in (47) is given by

$$\Pi_n(\gamma) = \sum_x p_n^x |x\rangle\langle x|^B \otimes (\rho_{n,x}^{\Lambda Q} - e^{n\gamma}\rho_n^{\Lambda Q}). \tag{50}$$

[1]According to the terminology of [10] it is the density matrix which one can associate to a *c-q resource* given by the ensemble $\{p_n^x, \rho_{n,x}^Q\}$.

Note that

$$\begin{aligned} & \mathrm{Tr}\Big[\{\Pi_n(\gamma)\geq 0\}\Pi_n(\gamma)\Big] \\ = \; & \mathrm{Tr}\Big[\{\Pi_n(\gamma)\geq 0\}\Big(\sum_{x\in\mathcal{X}^{(n)}} p_n^x|x\rangle\langle x|^B\otimes\big(\rho_{n,x}^{\Lambda Q}-e^{n\gamma}\rho_n^{\Lambda Q}\big)\Big] \\ = & \sum_x p_n^x \mathrm{Tr}\Big[\{\rho_{n,x}^{\Lambda Q}\geq e^{n\gamma}\rho_n^{\Lambda Q}\}\big(\rho_{n,x}^{\Lambda Q}-e^{n\gamma}\rho_n^{\Lambda Q}\big)\Big]. \end{aligned} \tag{51}$$

Hence, $\underline{S}(B:\Lambda Q)$ is equivalently given by

$$\sup\Big\{\gamma:\lim_{n\to\infty}\sum_x p_n^x\mathrm{Tr}\big[\{\rho_{n,x}^{\Lambda Q}\geq e^{n\gamma}\rho_n^{\Lambda Q}\}\big(\rho_{n,x}^{\Lambda Q}-e^{n\gamma}\rho_n^{\Lambda Q}\big)\big]=1\Big\}.$$

This implies that for any $\gamma<\underline{S}(B:\Lambda Q)$,

$$\lim_{n\to\infty}\sum_x p_n^x\mathrm{Tr}\big[\{\rho_{n,x}^{\Lambda Q}<e^{n\gamma}\rho_n^{\Lambda Q}\}\rho_{n,x}^{\Lambda Q}\big]=0. \tag{52}$$

For $M_n=\lceil e^{nR}\rceil$, Lemma 4 ensures the existence of a sequence of codes $\{\mathcal{C}^{(n)}\}_{n=1}^{\infty}$ of size $|\mathcal{C}^{(n)}|=\lceil e^{nR}\rceil$, such that for each n

$$\begin{aligned} P_e(\mathcal{C}^{(n)}) \;\leq\; & 2\sum_{x\in\mathcal{X}^{(n)}} p_n^x\mathrm{Tr}\big[\{\rho_{n,x}^{\Lambda Q}-e^{n\gamma}\rho_n^{\Lambda Q}\leq 0\}\rho_{n,x}^{\Lambda Q}\big] \\ & +4e^{-n\gamma}\lceil e^{nR}\rceil, \end{aligned} \tag{53}$$

for any $\gamma\in\mathbf{R}$ and $c>0$. From (52) it follows that for any $\gamma<\underline{S}(B:\Lambda Q)$, the first term on the RHS of (53) vanishes in the limit $n\to\infty$. For all $\delta>0$, there exists $n_0\in\mathbf{N}$, such that for all $n\geq n_0$, $\lceil e^{nR}\rceil\leq e^{n(R+\delta)}$. Hence,

$$4e^{-n\gamma}\lceil e^{nR}\rceil\leq 4^{-n(\gamma-(R+\delta))},$$

which vanishes as $n\to\infty$ for $\gamma>R+\delta$. Since δ is arbitrary, it follows that any rate $R<\gamma<\underline{S}(B:\Lambda Q)$ is achievable. More generally, any rate $0<R<\underline{S}(B:\Lambda Q)$ is achievable.

To prove the (weak) converse we are required only to show that for any code with rate larger than the capacity, there exists a probability distribution on the codewords such that the average probability of error does not vanish asymptotically.

Define a family of codes of size M_n by the average state of the codewords ρ_n^Q. Note that the family includes all possible sets of M_n codewords with the same average state. Given the family $\{M_n,\rho_n^Q\}_{n=1}^{\infty}$ we can extend ρ_n^Q to any separable state ρ^{AQ} on an enlarged Hilbert space. The outcome of any measurement on A is thus classically correlated with a state on Q.

Explicitly, we can assign the message that has been sent with the outcome of the set of measurements on A, described by a POVM $\{E_{n,i}^A\}$, such that message $i\in\{1,2,...,M_n\}$ is generated with probability

$$p_i=\mathrm{Tr}\big[(E_{n,i}^A\otimes I_n^Q)\rho_n^{AQ}\big]. \tag{54}$$

and results in the codeword

$$\rho_{n,i}^Q=\mathrm{Tr}_A\big[\sqrt{E_{n,i}^A\otimes I_n^Q}\,\rho_n^{AQ}\sqrt{E_{n,i}^A\otimes I_n^Q}\big] \tag{55}$$

which is then sent throught the noisy channel.

The average probability of error can thus be expressed as

$$\begin{aligned} P_e(\mathcal{C}^{(n)}) &= 1-\sum_{i=1}^{M_n}p_i\mathrm{Tr}\big[E_{n,i}^Q\Lambda_n^Q\rho_{n,i}^Q)\big] \\ &= 1-\sum_{i=1}^{M_n}\mathrm{Tr}\big[(E_{n,i}^A\otimes E_{n,i}^Q)\rho_n^{A\Lambda Q}\big], \end{aligned} \tag{56}$$

where $\rho_n^{A\Lambda Q}=(I_n^A\otimes\Lambda_n^Q)\rho_n^{AQ}$.

From Lemma 1 it then follows that

$$\begin{aligned} P_e(\mathcal{C}^{(n)}) &\geq 1-\mathrm{Tr}\big[\{\Pi_n(\gamma)\geq 0\}\Pi_n(\gamma)\big] \\ &\quad -e^{n\gamma}\mathrm{Tr}\big[\sum_{i=1}^{M_n}E_{n,i}^A\rho_n^A\otimes E_{n,i}^Q\rho_n^{\Lambda Q}\big] \\ &= 1-\mathrm{Tr}\big[\{\Pi_n(\gamma)\geq 0\}\Pi_n(\gamma)\big] \\ &\quad -e^{n\gamma}\sum_{i=1}^{M_n}p_i\mathrm{Tr}\big[E_{n,i}^Q\rho_n^{\Lambda Q}\big] \end{aligned} \tag{57}$$

with $\Pi_n(\gamma)=\rho_n^{A\Lambda Q}-e^{n\gamma}\rho_n^A\otimes\rho_n^{\Lambda Q}$, and where the probability p_i is given by (54).

Choosing only those POVMs such that

$$p_i=\mathrm{Tr}\big[E_{n,i}^A\rho_n^A\big]=\frac{1}{M_n} \tag{58}$$

is sufficient to show that any code of size M_n is not reliable. In this case

$$P_e(\mathcal{C}^{(n)})\geq 1-\mathrm{Tr}\big[\{\Pi_n(\gamma)\geq 0\}\Pi_n(\gamma)\big]-\frac{e^{n\gamma}}{M_n} \tag{59}$$

and for any $\delta>0$, choose $M_n=\lceil e^{nR}\rceil$ where $R=\underline{S}(A:\Lambda Q)+2\delta$, and $\gamma=\underline{S}(A:\Lambda Q)+\delta$. Thus, the third term on the RHS of (59) vanishes in the limit $n\to\infty$. However, the difference of the first two terms does not vanish and we have $\limsup_{n\to\infty}P_e(\mathcal{C}^{(n)})\geq\epsilon_0$ for some $\epsilon_0>0$.

We thus conclude that the classical capacity of a sequence of channels $\Lambda=\{\Lambda_n^Q\}_{n=1}^{\infty}$ is given by

$$C(\Lambda)=\max_{\rho^{BQ}\in\mathcal{Q}}\underline{S}(B:\Lambda Q) \tag{60}$$

where $\mathcal{Q}$ denotes the set of sequences of classical–quantum states in $\mathcal{B}(\mathcal{H}_{BQ})$, with $\mathcal{H}_{BQ}$ being a sequence of Hilbert spaces $\mathcal{H}_{BQ}:=\{\mathcal{H}_B^{(n)}\otimes\mathcal{K}_Q^{(n)}\}_{n=1}^{\infty}$, The monotonicity of the inf-spectral mutual information rate under CPTP maps (see [1]) implies that $\underline{S}(B:\Lambda Q)\equiv\underline{S}(AA':\Lambda Q)\geq\underline{S}(A:\Lambda Q)$. This ensures that optimization over classical-quantum states is equivalent to optimization over separable states, thus yielding the statement 45 of Theorem 4.

V. Dense Coding

Dense coding is the protocol by which prior shared entanglement between a sender (Alice) and a receiver (BOB) is exploited for sending classical messages through a noiseless quantum channel. Let $\rho_n^{AB}\in\mathcal{H}_A^{(n)}\otimes\mathcal{H}_B^{(n)}$ be an entangled mixed state that Alice and Bob initially share. As in Section IV, Alice has a set of messages, labelled by the elements of the set $\mathcal{M}_n=\{1,2,\ldots,M_n\}$, which she wishes to communicate to Bob. However, the quantum channel that she uses is noiseless. She encodes her messages into her part, A, of the bipartite system AB which is in the state ρ_n^{AB}. The codewords are given by

$$\phi_n(i) := \rho_{n,i}^{AB} = (\mathcal{E}_{n,i}^A \otimes \mathrm{id}^B)\rho_n^{AB},$$

for $i = \mathcal{M}_n$. Here ϕ_n denotes the encoding map for a code of size M_n as defined in terms of the CPTP maps $\mathcal{E}_{n,i}^A$, $i \in \mathcal{M}_n$. Let Bob's measurement on the states $\rho_{n,i}^{AB}$ that he receives, be given by $E_n^{AB} = \{E_{n,i}^{AB}\}_{i=1}^{M_n}$, with each $E_{n,i}^{AB} \geq 0$ and $\sum_{i=1}^{M_n} E_{n,i}^{AB} \leq I_n^{AB}$. The average probability of error of the code $\mathcal{C}^{(n)} = (M_n, \phi_n^A, E_n^{AB})$ is given by

$$P_e(\mathcal{C}^{(n)}) := \frac{1}{M_n}\sum_{i=1}^{M_n}\left(1 - \mathrm{Tr}(\rho_{n,i}^{AB}E_{n,i}^{AB})\right), \tag{61}$$

The dense coding capacity for a sequence of bipartite states $\rho^{AB} = \{\rho_n^{AB}\}_{n=1}^\infty$ is defined as

$$C_{DC} := \sup R, \tag{62}$$

where R is an achievable rate.

Theorem 5: The dense coding capacity for a sequence of bipartite states $\rho^{AB} = \{\rho_n^{AB}\}_{n=1}^\infty$ is given by

$$C_{DC} = \log d - \min_\Lambda \overline{S}(\Lambda A|B) \tag{63}$$

where $\Lambda = \{\Lambda_n^A\}_{n=1}^\infty$ is a sequence of CPTP maps on A.

Proof: [Converse] For a code $\mathcal{C}^{(n)}$ of M_n codewords $\rho_{n,i}^{AB} = (\phi_{n,i}^A \otimes \mathrm{id}^B)\rho_n^{AB}$, and measurement operators $E_{n,i}^{AB}$, $i = 1, \ldots, M_n$, the average probability of error (61) satisfies

$$\begin{aligned} P_e(\mathcal{C}^{(n)}) &\geq 1 - \frac{1}{M_n}\sum_i \mathrm{Tr}\big[E_{n,i}^{AB}\,\rho_{n,i}^{AB} - e^{-n\gamma}I_n^A \otimes \rho_n^B\big] \\ &\quad - \frac{e^{-n\gamma}}{M_n}\mathrm{Tr}\Big[E_{n,i}^{AB}\,(I_n^A \otimes \rho_n^B)\Big] \\ &\geq 1 - \frac{1}{M_n}\sum_i \mathrm{Tr}\big[\Pi_{n,i}(\gamma)\big] - \frac{e^{-n\gamma}}{M_n}\mathrm{Tr}\, I_n^A \\ &\geq 1 - \max_i \mathrm{Tr}\big[\Pi_{n,i}(\gamma)\big] - \frac{e^{n(\log d - \gamma)}}{M_n} \end{aligned} \tag{64}$$

where $\Pi_n^i(\gamma) = \{\rho_{n,i}^{AB} \geq e^{-n\gamma}I_n^A \otimes \rho_n^B\}\big(\rho_{n,i}^{AB} - e^{-n\gamma}I_n^A \otimes \rho_n^B\big)$. In the above we have used Lemma 1 and the facts that $\sum_i E_{n,i}^{AB} \leq I_n^{AB}$ and $\mathrm{Tr}\, I_n^A = e^{n\log d}$.

If we then assume that $M_n \geq e^{nR} = \log d - \min_\Lambda \overline{S}(\Lambda A|B) + 2\delta$ for some $\delta > 0$, then we can choose $\gamma = \min_\phi \overline{S}(\phi A|B) - \delta$, and we find

$$\limsup_{n\to\infty} P_e(\mathcal{C}^{(n)}) \geq \epsilon_0 > 0 \tag{65}$$

implying $C_{DC} \leq \log d - \min_\phi \overline{S}(\phi A|B)$. ∎

Proof: [Coding] Lemma 4, adapted to the case of dense coding, states that for any $n \in \mathbf{N}$, $M \in \mathbf{N}$, and $\gamma \in \mathbf{R}$, given a probability distribution $\{p_n^x\}$ on $\mathcal{X}^{(n)}$, where $\mathcal{X}^{(n)}$ is a finite set of indices, there exists a code $\mathcal{C}^{(n)}$ of size $|\mathcal{C}^{(n)}| = M$, whose average probability of error satisfies the bound

$$\begin{aligned} P_e(\mathcal{C}^{(n)}) \;\leq\; & 2\sum_{x\in\mathcal{X}^{(n)}} p_n^x \mathrm{Tr}\big[\{\rho_{n,x}^{AB} < e^{n\gamma}\rho_n^{AB}\}\rho_{n,x}^{AB}\big] \\ & + 4e^{-n\gamma}M, \end{aligned} \tag{66}$$

where

$$\rho_n^{AB} := \sum_{x\in\mathcal{X}^{(n)}} p_n^x \rho_{n,x}^{AB}.$$

Choose $\mathcal{X}^{(n)}$ to be a set of size $N_n = d^{2n}$ and define a probability distribution $\{p_n^x\}$ on it, with $p_n^x = 1/N_n = e^{-2n\log d}$ for each $x \in \mathcal{X}^{(n)}$. Further, consider states $\rho_{n,x}^{AB}$ defined as follows:

$$\rho_{n,x}^{AB} := \big(\mathcal{U}_{n,x}^A\Lambda_n^A \otimes \mathrm{id}^B)\big)\rho_n^{AB}.$$

Here Λ_n^A denote quantum operations for which the sequence $\{\Lambda_n^A\}_{n=1}^\infty$ minimizes $\overline{S}(\Lambda A|B)$, and (ii) $\mathcal{U}_{n,x}^A$, $x \in \mathcal{X}^{(n)}$, denotes unitary encodings with the shift-multiply operators $U_{(p,q)}$, with $p, q \in \{0, 1, \ldots, (d^n - 1)\}$, which are defined as follows ([7], [12]):

$$U_{(p,q)}|j\rangle = e^{\frac{2\pi pj}{d}}|j + q\,(\mathrm{mod}\,d)\rangle,$$

with $\{|j\rangle : j \in \{0, 1, \ldots, (d^n - 1)\}$ being an orthonormal basis in a d^n-dimensional Hilbert space.

Let

$$\rho_n^{\Lambda AB} := (\Lambda_n^A \otimes \mathrm{id}^B)\rho_n^{AB}.$$

For the ensemble $\{p_n^x, \rho_{x,n}^{AB}\}$

$$\begin{aligned} \sum_{x\in\mathcal{X}^{(n)}} p_n^x\rho_{n,x}^{AB} &= \sum_{x\in\mathcal{X}^{(n)}} p_n^x\big(\mathcal{U}_{n,x}^A \otimes \mathrm{id}^B)\rho_n^{\Lambda AB} \\ &= \frac{I_n^A}{d^n} \otimes \rho_n^B, \end{aligned} \tag{67}$$

where ρ_n^B is the reduced density matrix of the state $\rho_n^{\Lambda AB}$.

For the ensemble $\{p_n^x, \rho_{x,n}^{AB}\}$ defined above, let

$$\alpha_n := \sum_{x\in\mathcal{X}^{(n)}} p_n^x \mathrm{Tr}\big[\{\rho_{n,x}^{AB} \geq e^{n\gamma}\rho_n^{AB}\}\rho_{n,x}^{AB}\big]$$

We have that

$$\begin{aligned} \alpha_n &\geq \frac{1}{N_n}\sum_{x\in\mathcal{X}^{(n)}} \mathrm{Tr}\big[\{\rho_{n,x}^{AB} \geq e^{n\gamma}\rho_n^{AB}\} \\ &\qquad \times \big(\rho_{n,x}^{AB} - e^{n\gamma}\rho_n^{AB}\big)\big] \\ &= \mathrm{Tr}\big[\{\rho_n^{\Lambda AB} \geq e^{-n(\log d - \gamma)}I_n^A \otimes \rho_n^B\} \\ &\qquad \times \big(\rho_n^{\Lambda AB} - e^{-n(\log d - \gamma)}I_n^A \otimes \rho_n^B\big)\big]. \end{aligned} \tag{68}$$

In the above we have made use of the fact that the trace remains invariant under a unitary transformation. If $\gamma = \log d - \overline{S}(\Lambda A|B) - \delta$ for any $\delta > 0$, the RHS of (68) goes to one as $n \to \infty$. Hence the RHS of (66) vanishes asymptotically, implying that a rate $R = \log d - \min_\Lambda \overline{S}(\Lambda A|B) - \delta$ is achievable for any $\delta > 0$. ∎

A. Reduction to the i.i.d. Case

For entanglement resources which are tensor products of identical bipartite states $\rho_N^{AB} = \varrho_{AB}^{\otimes N}$, with $\varrho_{AB} \in \mathcal{B}(\mathcal{H})$, the dense coding capacity was shown in [11] to be given by

$$C_{DC} = \log d + S(B) - \inf_N \inf_{\Lambda_A^{(N)}} \frac{1}{N} S\big((\Lambda_A^{(N)} \otimes \mathrm{id}_B^{(N)})\varrho_{AB}^{\otimes N}\big). \tag{69}$$

Here $S(B) = S(\varrho_B)$, where ϱ_B is the reduced density matrix of the system B, corresponding to the state ϱ_{AB}.

For sequences of i.i.d. states $\omega = \{\vartheta^{\otimes n}\}_{n=1}^{\infty}$ and $\sigma = \{\varsigma^{\otimes n}\}_{n=1}^{\infty}$, Theorem 4 of [5] states that

$$\underline{D}(\omega\|\sigma) = D(\vartheta\|\varsigma) = \overline{D}(\omega\|\sigma). \tag{70}$$

For bipartite states $\vartheta = \vartheta_{AB}$ and $\varsigma = \frac{1}{d} I_A \otimes \vartheta_B$, (70) implies that

$$\overline{S}(A|B) = S(A|B) = \underline{S}(A|B), \tag{71}$$

where $S(A|B) = S(\vartheta^{AB}) - S(\vartheta^B)$. This is because $\log d - \overline{S}(A|B) = \underline{D}(\omega\|\sigma) = D(\vartheta\|\varsigma) = \log d - \frac{1}{n} S\big(\vartheta_{AB}^{\otimes n} | \vartheta_B^{\otimes n}\big) = \log d - S(A|B)$, and similarly for $\overline{D}(\omega\|\sigma)$. If instead, we choose ϑ_{AB} and ς_{AB} to be states in $\mathcal{B}(\mathcal{H}^{\otimes N})$, given by

$$\vartheta_{AB} := \big(\Lambda_A^{(N)} \otimes \mathrm{id}_B^{(N)}\big) \varrho_{AB}^{\otimes N},$$

and

$$\varsigma_{AB} = \tfrac{1}{d^N} I_A^{(N)} \otimes \varrho_B^{\otimes N},$$

then (70) yields the identity

$$\overline{S}(\Lambda^{(N)} A|B) = \frac{1}{N} S\Big((\Lambda_A^{(N)} \otimes \mathrm{id}_B^N) \varrho_{AB}^{\otimes N} \Big) - S(\varrho_B). \tag{72}$$

Hence, in this case our expression (63) for the dense coding capacity reduces to (69).

References

[1] G. Bowen and N. Datta, "Beyond i.i.d. in Quantum Information Theory," *quant-ph/0604013, Proceedings of the 2006 IEEE International Symposium on Information Theory*.

[2] S. Verdu and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147–1157, 1994.

[3] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer-Verlag, 2002.

[4] T. Ogawa and H. Nagaoka, "Strong converse and stein's lemma in quantum hypothesis testing," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2428–2433, 2000.

[5] H. Nagaoka and M. Hayashi, "An information-spectrum approach to classical and quantum hypothesis testing for simple hypotheses," *quant-ph/0206185*, 2002.

[6] M. Hayashi and H. Nagaoka, "General formulas for capacity of classical–quantum channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1753–1768, 2003.

[7] T. Hiroshima, "Optimal dense coding with mixed state entanglement," *quant-ph/0009048*

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[9] H. Araki and E. Lieb, "Entropy inequalities," *Comm. Math. Phys.*, vol. 18, pp. 160–170, 1970.

[10] I. Devetak and A. Winter, "Distilling common randomness from bipartite quantum states," *quant-ph/0304196*

[11] M. Horodecki *et al.*, "Classical capacity of a noiseless channel assisted by noisy entanglement," *Quantum Information and Computation*, vol. 1, No. 3, pp. 70-78, 2001., *quant-ph/0106080*

[12] G. Bowen, "Classical information capacity of superdense coding," *Phys. Rev. A*, vol. 63, 022302, 2001.